

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number
WO 01/27809 A2

(51) International Patent Classification⁷: **G06F 17/30**

(21) International Application Number: **PCT/NL00/00742**

(22) International Filing Date: 16 October 2000 (16.10.2000)

(25) Filing Language: Dutch

(26) Publication Language: English

(30) Priority Data:
1013297 15 October 1999 (15.10.1999) NL

(71) Applicant (*for all designated States except US*): PLANT RESEARCH INTERNATIONAL B.V. [NL/NL]; Drovendaalsesteeg 1, NL-6708 PB Wageningen (NL).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): VAN DER KRIEKEN, Wilhelmus, Maria [NL/NL]; Laantje van Anton Pieck 2, NL-6708 RE Wageningen (NL). KODDE, Jan [NL/NL]; Van Doesburglaan 28, NL-6708 MC Wageningen (NL).

(74) Agents: LAND, Addick, Adrianus, Gosling et al.; Arnold & Siedsma, Sweelinckplein 1, NL-2517 GK Den Haag (NL).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A2
WO 01/27809

(54) Title: VISUALIZING RELATIONS IN DATA SETS

(57) Abstract: The present invention relates to a method for comparing and/or analyzing data files, such as for instance are obtained in one or more array forms or in a matrix form, wherein: the data and relations from the arrays or matrices with mutual correlations are placed via rearrangement in a virtual matrix, whereby these mutual relations or correlations can be made easily visible for a user, for instance on a screen.

VISUALIZING RELATIONS IN DATA SETS

Introduction

Comparison of data files shows whether there are similarities or differences in the files. Generally the mutual relation between two sets of data is compared, and sometimes between a number of sets. This comparison 5 is often a time-consuming activity ("data-mining") wherein errors can easily be made. The invention described here relates to a technique for comparing two or more data sets by showing the differences and/or similarities between the sets in figures. This technique 10 is for instance important for comparing data from experiments with micro-arrays, although other data sets can also be compared with this technique. Other examples are data sets from:
cosmology,
15 mathematics,
screening of the population,
patient screening,
sociological research,
physical determinations,
20 biotechnology,
High throughput systems.

Analysis of data files of DNA micro-arrays (and DNA chips) are further elaborated here. Files of DNA-micro-arrays are an example of biotechnological data 25 files (other examples hereof are for instance files of dot blots, cDNA-AFLP, Northern blots, Southern blots and protein-arrays from proteomics). The use of DNA micro-arrays is growing enormously in molecular genetics. In this technique DNA fragments associated with particular 30 genes are placed as little "spots" on a glass plate (the sequence of the DNA fragments and function of the associated gene can be known after analysis). With particular equipment it is possible to make many copies of such a glass plate (the position of the DNA spots on 35 the glass plate is the same in all copies.) Many

thousands of DNA (or cDNA or RNA) spots can be placed on a small glass plate (of for instance one square inch). The DNA, cDNA or RNA in the spots on the array are further designated as "targets". With micro-arrays it is

5 possible to show via determined standard DNA-techniques (hybridization techniques with labeled probes) which genes are active in a determined tissue, under particular conditions. In brief, this proceeds as follows: the DNA targets on the array are hybridized (compounded) with

10 DNA, RNA or cDNA originating from the tissue for testing. This DNA, RNA or cDNA is marked with a particular colour (for instance yellow) and is therefore referred to as probe. After hybridization with the probe the intensity of the colouring is proportional to the (relative)

15 expression level of the gene. Two DNA or RNA probes can also be tested simultaneously. The second probe originates for instance from tissue which has undergone a treatment other than the tissue from which the first probe is made. The second probe can be marked with a

20 different colour (for instance red). Targets on the micro-array which are yellow or red after the hybridization are only expressed in one of the tissues and targets with an intermediate colour (orange-like colours) are expressed in both tissues. With optical

25 equipment the colour of a determined substance can be measured very precisely and the mutual expression difference in the treated and untreated tissue can hereby be determined per target. The differences in expression can be easily recorded: genes which are expressed

30 specifically enough under particular conditions (for instance from tissue after administering of a determined medicine) can be isolated and then studied further. There are different variants for using DNA micro-arrays (cDNA micro-arrays, subtractive cDNA micro-arrays, micro-arrays

35 wherein EST, expressed sequence tags, are spotted on the glass plate as targets etc.). Because large numbers of genes are compared (often many thousands per array and these numbers are becoming ever greater) large data files

are created. These data files have heretofore been processed in rather unsophisticated manner. Per target the change in expression in different conditions is analyzed and stored (via a spreadsheet, via a coloured representation of the array, wherein the colour represents the change in gene expression or representation in a "tree diagram"). Because this method of comparison is very time-consuming (the many thousands of targets must be mutually compared) and because errors can easily occur, in practice a pre-formulated expression profile is often sought and the strongest exponent hereof is selected, resulting in the loss of much information. In such a one-dimensional analysis the mutual relations between the different expression patterns therefore often do not come to light very well, while this is precisely the valuable information which can be generated with micro-arrays. Even if only one specific expression profile is being sought, it is very useful to know whether this profile can be placed in a larger context. It may for instance be important to know whether genes with a specific expression pattern are involved in a plurality of processes. This also reduces the chance of the found profile being the result of measurement errors.

Object of the data analysis program specified

here

The object of the analysis program described here is to visualize expression data obtained using micro-arrays (or data from other data files) in the form of virtual arrays. The array elements are herein rearranged in a manner such that genes with a comparable expression pattern are grouped together at a meaningful position in the virtual array. Mutual relations hereby become visible in rapid, ordered and clear manner.

This is achieved by rearranging the targets of the original array on a virtual array. The targets associated with a treatment with a determined specificity (for instance specific gene expression after administering of a medicine) are herein placed together

4

in a figure representing a virtual array. The targets which are expressed in another treatment are also placed together, and overlaps within these groups of genes are also visualized. Relations in gene expression in 5 different treatments hereby become immediately apparent. Genes which are always expressed together and form, as it were, networks of gene expression are also identified. The rearrangement is such that the information about the original position of the targets is saved so that it is 10 possible to retrieve which gene (function and sequence) is specifically expressed. The new location of the target in this virtual array is determined by the relative expression level relative to the expression of the control genes or of genes which are expressed in other 15 treatments. In a single (1 determination/1 physiological condition) experiment the results are easy to compare. If however a plurality of mutually related conditions (for instance a time array and/or a concentration array) have to be compared, the location of the target in the virtual 20 array also depends on the result of the other measurements within the inputted array.

Further advantages, features and details of the present invention will be elucidated on the basis of the following description with reference to the annexed 25 drawings, in which:

fig. 1 shows a diagram of a regeneration process in a culture medium as embodiment of the present invention,

fig. 2 shows a graphic representation of the 30 result of the experiment according to fig. 1,

fig. 3 is a graphic representation of regrouping of the representation of fig. 2,

fig. 4 is a graphic representation of the groups of genes derived from fig. 3 which are expressed,

35 fig. 5 is a graphic representation of a regrouping obtained according to method 1 below, and

fig. 6 is a graphic representation of a regrouping obtained according to method 2 below.

The foregoing is illustrated with an experiment. Administering of the plant hormones auxin and cytokinin to plant tissue in vitro results in regeneration of roots, shoots and/or white or green callus. These regeneration processes are dependent on the concentration of auxin and cytokinin in the culture medium (Figure 1). Gene expression values associated with an array tested under the conditions as shown in figure 1 are shown graphically in figure 2. The colour indicates, between two threshold values, the height of the expression: (blue is low and red is high expression). The groups of genes which are expressed in relation to a determined treatment are shown grouped together in figure 3. The function for the genes which can be derived herefrom is shown in figure 4. The set-up of the physiological experiment in combination with the rearrangement of the targets on the virtual arrays determines the information which can be obtained. What these figures clearly illustrate is that rearrangement of the results in this manner gives a much clearer picture of possible relations between the genes. Analysis without the technique described here will take up very much time and it is doubtful whether the relations described here are found, since this will depend on the time available and the experience of the researcher performing the analyses. In this experiment only genes are shown whose expression increases relative to gene expression in the untreated tissue. In similar manner, genes which are expressed less relative to this tissue can also be processed in the virtual array. This rearrangement can be easily automated by means of a computer program.

Methods

Shown in figure 1 is a typical example of an experiment wherein gene expression, by means of micro-arrays, is studied in relation to different physiological conditions. Within this data-analysis method virtual arrays are made by ordering individual targets on the arrays such that they correspond as much as possible with

the physiological incubation conditions in the experiment. Hereby the position of a particular target on the virtual array is hereby no longer random but on the contrary correlated to its function at a determined 5 physiological condition in the experiment. Fig. 3 shows the result of an ideal virtual arrangement of the results of fig. 1.

In order to obtain a good arrangement two algorithms for the rearrangement are shown below. A 10 method wherein clusters are generated with targets which have mutually comparable gene expression patterns in tissues incubated under determined conditions. In the second method the targets are rearranged on the basis of their individual expression properties under such 15 conditions.

The algorithms are capable of (virtually) ordering the results obtained in experiments with micro-arrays. The micro-arrays consist of DNA targets having therein copies of complete genes or parts thereof. In the 20 experiments the expression is measured of the genes which are present on the micro-array. In order to measure this expression a known mRNA pool, isolated from tissue, is converted into labeled cDNA. The labeled cDNAs can bind to complementary DNA in the targets. The expression level 25 of a gene determines its proportion in the mRNA pool. This expression level can thus be measured by determining the quantity of label (of the cDNA) on a particular target.

Taken as starting point in the experiments are 30 tissues incubated under different conditions. These incubations are such that they have a determined mutual relation (for instance a concentration array of a determined substance and/or a time array). The example shows an experiment with two substances (auxin and 35 cytokinin). The function and relation of the targets can be determined from the obtained virtual arrangement.

The data are expression levels which are measured on the micro-arrays (quantity of one or more

7
labels on a target) in the above stated experiments. For the algorithm this data is placed in columns (or rows). Each column (or row) herein represents the data obtained at a determined incubation condition (determined 5 concentration of auxin or cytokinin). A row (or column) represents the expression results of a target under all incubation conditions.

Method 1. Clustering of genes with comparable expression pattern on virtual arrays.

10 **Data transformation**

When the data is sorted, the expression values are first classified in a limited number of levels. Each level therefore represents an expression value between two preselected limit values. In those targets where the 15 expression value lies below the lowest limit value, the level value zero is assigned (the value zero is assigned to low expression levels). The final outcome can be optimized by choosing the limit values of the different levels. If down-regulation of genes is also being 20 studied, the zero value can for instance be assigned to another level.

Sorting of the data on the basis of clusters.

The targets in the different columns are clustered. This can be done for instance as follows: A 25 binary code is assigned to each target. One binary number is assigned (0 or 1) to each expression value of a determined column. One of the expression levels, for instance expression level 0, is given the binary number 0 and all other expression levels the value 1. A row with 30 numbers thus results which forms the binary code. All targets with the same binary code are clustered.

Placing of arrays during virtual arrangement.

Prior to the virtual arrangement, the arrays representing the different incubation conditions are 35 placed in a determined sequence/position. This placing represents for instance the test set-up. In this case the incubation with the lowest concentration of substance 1 is placed on the left and that with the highest on the

right (and at the bottom the lowest concentration of substance 2 and at the top the highest concentration of substance 2 (as is usual in the plotting of figures).

Calculation of the optimal position of the 5 clusters on the virtual arrays.

The virtual arrangement of clusters within the arrays can be based on the following. All targets within a cluster have the same, calculated position. If a cluster is expressed at a high concentration of substance 10 1 and of substance 2, this cluster is then also placed on the right at the top within all virtual arrays. In other words, the "X and Y" coordinates of the cluster (and the targets within this cluster) are dependent on the expression pattern of the clustered targets in the 15 different arrays.

Placing of the clustered targets on the virtual arrays.

Once the X and Y coordinates have been calculated targets can be placed on the virtual arrays.

20 This takes place cluster by cluster (all targets within a cluster have in the first instance the same X and Y coordinates). The sequence in which clusters are placed can be in order of increasing or decreasing X and/or Y value, in accordance with increasing cluster size, and so 25 on. If during placing the calculated position is already occupied, the closest empty position is chosen. The visualizing will be clearer by beginning with the target which has the highest total expression within a cluster.

If during placing of a new cluster the 30 calculated position is occupied, another position is chosen. This position can for instance be the closest position which is still available, although other placing strategies can also be envisaged. All targets within the cluster then acquire the X and Y coordinates of this new 35 position.

For a good overview of the virtual arrays (good delimitation of different clusters), it can be advantageous to make the virtual arrays larger than the

original arrays. In this manner an equal number of virtual and original arrays is generated. The expression values on the virtual arrays are herein visualized by means of colours, grey tones or hatchings which represent 5 the expression level. In addition, the virtual target also contains information about the original target (specification of the DNA, location on the original arrays etc).

The clustering can be based on more than two 10 dimensions, the visualizing (placing of the targets and clusters) is two-dimensional and the expression level can be shown as third dimension (via colour or hatching etc).

Method 2. Placing of individual targets in virtual arrays so that relations in gene expression 15 become apparent in relation to the set-up of the performed experiments.

Sorting and placing of individual targets on the virtual array.

This form of arrangement is based on the 20 calculation of an individual X and Y coordinate for each target in the virtual array. This calculation can for instance be done as follows: of each target the expression on a determined array is multiplied by the position of this array relative to the other arrays in 25 the physiological experiment in the X or respectively Y direction (in two-dimensional analysis, in the case of more dimensions calculations are of course performed with proportionally more coordinates). The expression of a determined target of figure 1 which is placed in the 30 physiological set-up on the left at the bottom is for instance multiplied by 1 to calculate both the x and the y coordinate. The expression of this target in the array to the right thereof is multiplied by 2 for the x coordinate and by 1 for the y coordinate. The expression 35 in the physiological condition hereabove is multiplied by 2 for both the x and the y coordinate. This is calculated for the target for all incubation conditions in the physiological experiment. All outcomes for a determined

target are then added together. Finally, this number is divided by the sum of all expression levels of the target in all incubation conditions. This calculation is performed in order to determine the x and the y position 5 of each target on the virtual array.

Form of the virtual array.

The form of the original array depends on the application onto the glass plate. The form of this array (number of rows and columns) may hereby not match the 10 physiological incubation conditions very well. It may therefore be advantageous to adapt the form of the virtual array to the performed experiments (see for instance figure 1).

Sorting of the data

15 The targets are sorted on the basis of these X and Y coordinates. The form (number of columns with targets or number of rows with targets) of the virtual array is taken into account here. If for instance the number of columns (equals the width of the array) is 20 taken into account, the targets are first sorted on the basis of increasing or decreasing y coordinates. The sorted targets are then subdivided into groups the size of the array width. Sorting within these groups subsequently takes place on the basis of increasing or 25 decreasing y coordinate.

Placing of targets on the virtual array

The sorted targets are placed in the virtual array row by row or column by column. The starting position (which of the four corners) and manner of 30 placing (rows or columns) depends on the sequence (the x or y coordinate first) and method (increasing or decreasing) of sorting. If a start is made with the x coordinate, the rearrangement is then particularly optimized, with the algorithm used here, for the variable 35 which is plotted on the x-axis. If a start is made with the y coordinate, rearrangement is then particularly optimized for the variable which is plotted on the y-axis. With other sorting and placing methods this may be

different again. In all cases a skilled researcher can analyse the data quickly and in orderly and reliable manner.

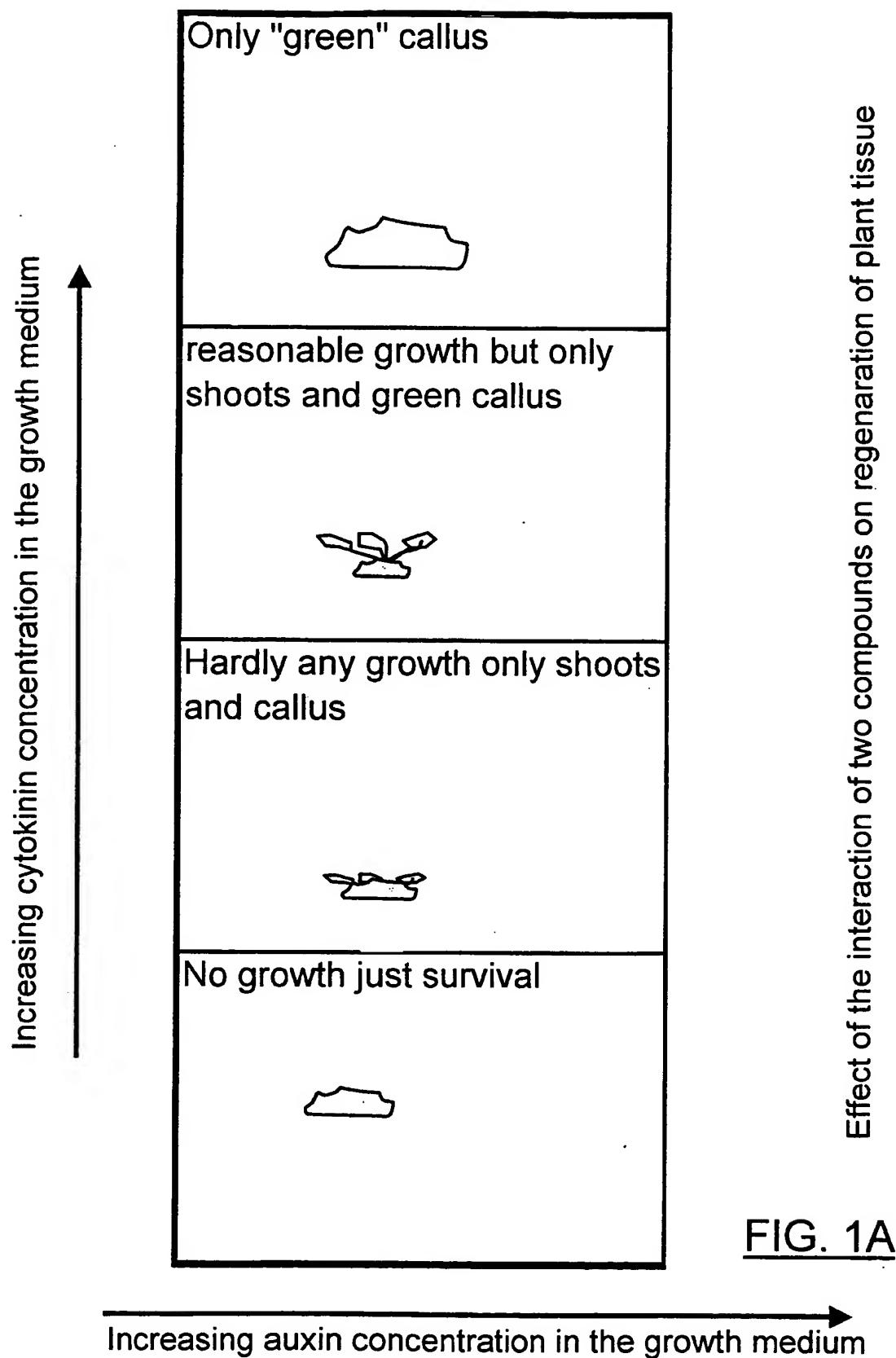
During placing the same number of virtual arrays is in principle generated as there are incubation conditions (real arrays) in the experiment. In the virtual array only the position of a target is changed. The change in the position is the same over all arrays (If for instance target with number 321 is placed in the top left-hand corner in array 1, this will also occur in all other arrays). The expression does not change but can be represented in a different manner. It is even recommended to represent the expression by means of colours (wherein a determined colour represents a determined expression range).

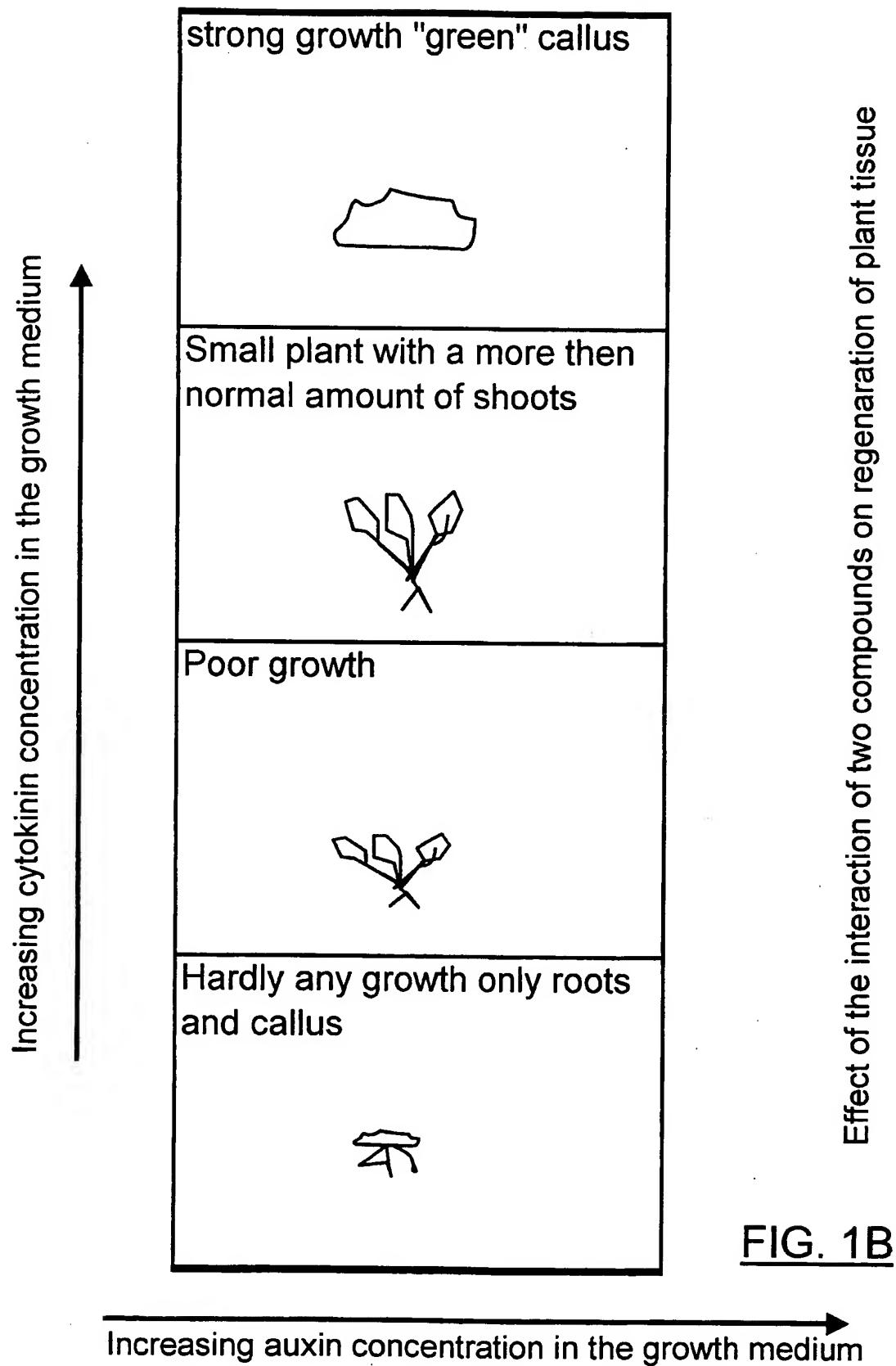
Interpolations or extrapolations for making virtual arrays.

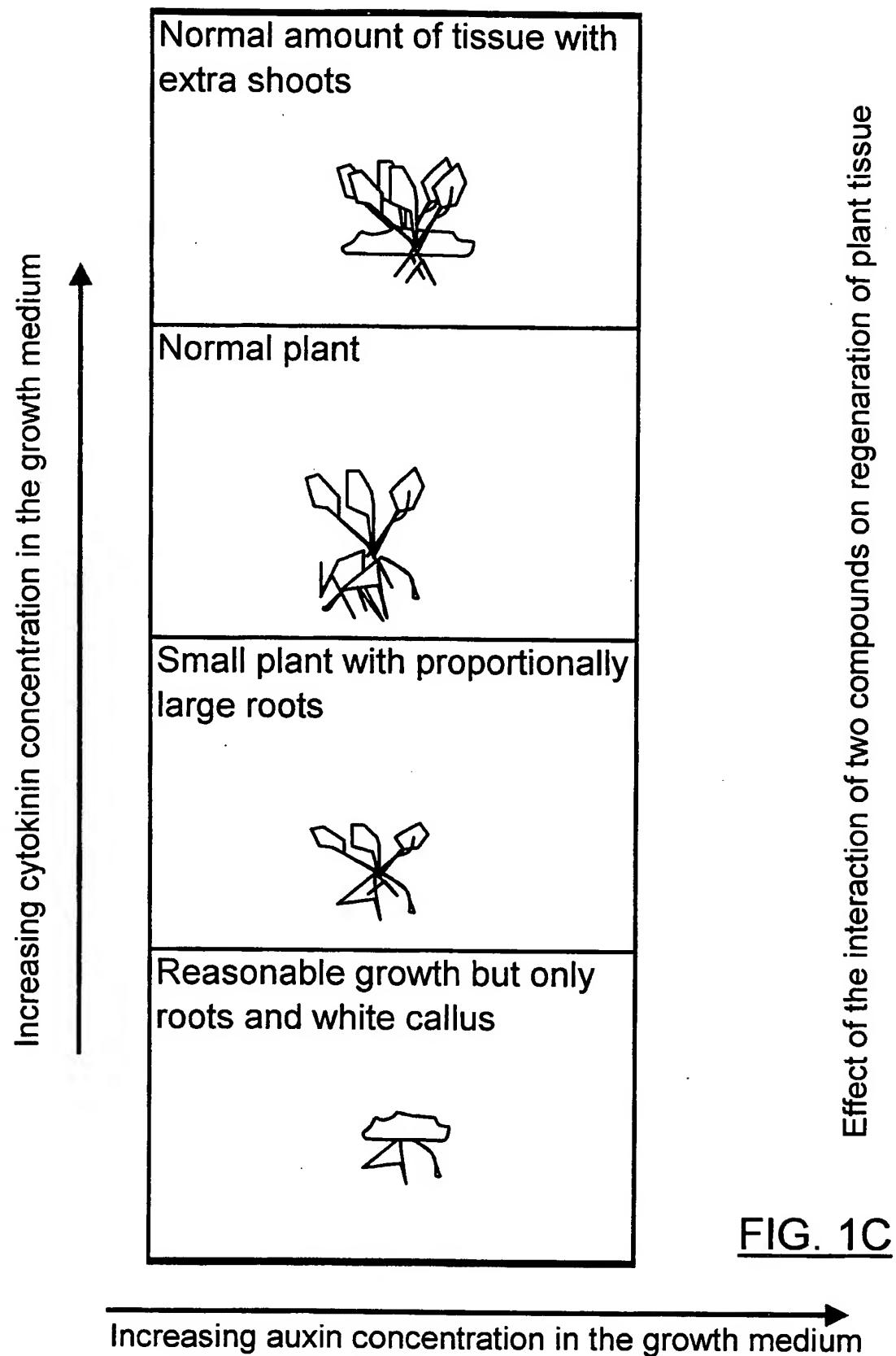
It is the case for both method 1 and 2 that it is possible via interpolation or extrapolation to make an array which can represent a determined, non-tested physiological condition. The expression data of a determined target are herein calculated via interpolation or extrapolation of expression levels which have been measured (associated with incubation conditions which are related to those of the calculated array).

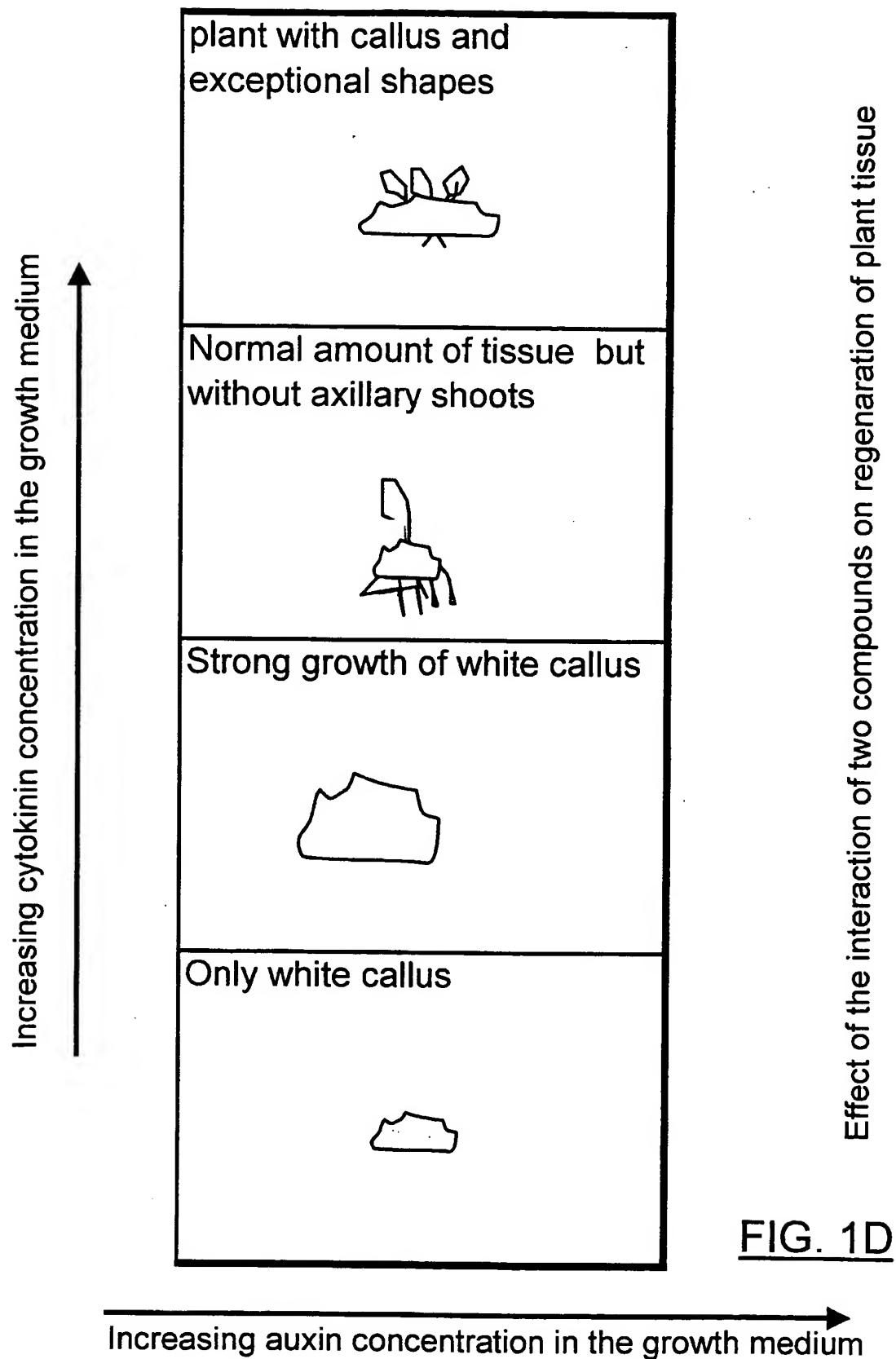
CLAIMS

1. Method for comparing and/or analyzing data files, such as for instance are obtained in one or more array forms or in a matrix form, wherein:
 - 5 - the data and relations from the arrays or matrices with mutual correlations are placed via rearrangement in a virtual matrix, whereby these mutual relations or correlations can be made easily visible for a user, for instance on a screen.
 - 10 2. Method as claimed in claim 1, wherein the data files originate from DNA micro-arrays or DNA chips.
 - 15 3. Method as claimed in claim 2, wherein expression values originating from the DNA micro-arrays are classified in a limited number of levels, whereby a binary code is preferably obtained.
 - 20 4. Method as claimed in claim 3, wherein targets are clustered in different columns of the micro-array.
 - 25 5. Method as claimed in any of the foregoing claims, wherein prior to the virtual arrangement the arrays are placed in a determined sequence in accordance with different incubation conditions.
 - 30 6. Method as claimed in claim 2, wherein at least one individual x and y coordinate is calculated for each target in the virtual array.
 - 35 7. Method as claimed in claim 6, wherein the expression of a determined array is multiplied by the position of this array relative to the preceding arrays.
 - 40 8. Method as claimed in claim 7, wherein targets are sorted on the basis of a first (x) respectively a second (Y) coordinate and the sorted targets are placed row by row, column by column, in the virtual array.
 - 45 9. Computer system, in which is stored a program for performing one or more of the above methods.
 - 50 10. Data file obtained according to the method from one or more of the above claims.









Increasing cytokinin concentration in the growth medium

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	
189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	
271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	290	
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	
331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	
391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	
421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	
451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	
511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	
541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	
571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	590	
591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	610	
619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	
649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	
679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709
709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	
739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	
769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	
799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	
829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859
859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889
889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919
919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949
949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979
979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999

Microarray data before rearrangement.

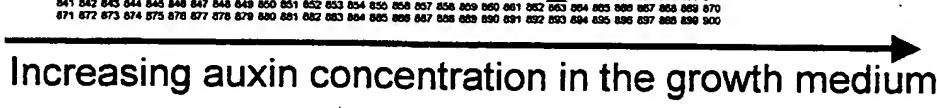


FIG. 2A

Increasing cytokinin concentration in the growth medium



Increasing auxin concentration in the growth medium

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270
271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330
331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390
391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420
421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450
451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510
511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540
541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570
571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600
601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630
631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660
661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690
691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720
721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750
751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780
781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810
811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840
841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870
871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900

FIG. 2B

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270
271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330
331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390
391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420
421	422	423	424	425	426	427	428	429	430	431	432																		

Increasing cytokinin concentration in the growth medium

Microarray data before rearrangement

Increasing auxin concentration in the growth medium

FIG. 2C

Increasing cytokinin concentration in the growth medium ↑

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115
122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146
153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177
189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213
214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238
239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263
264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325
331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385
386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410
412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436
445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469
473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497
498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522
523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547
548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572
573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597
598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622
623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647
648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672
673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697
698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722
723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747
751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775
781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805
811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835
841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865
871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895

Microarray data before rearrangement.

Increasing auxin concentration in the growth medium →

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115
122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146
153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177
189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213
214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238
239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263
264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325
331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385
391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415
416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440
451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505
512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536
542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566
572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596
597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621
631	632	633	634	635	636	637	638	6																

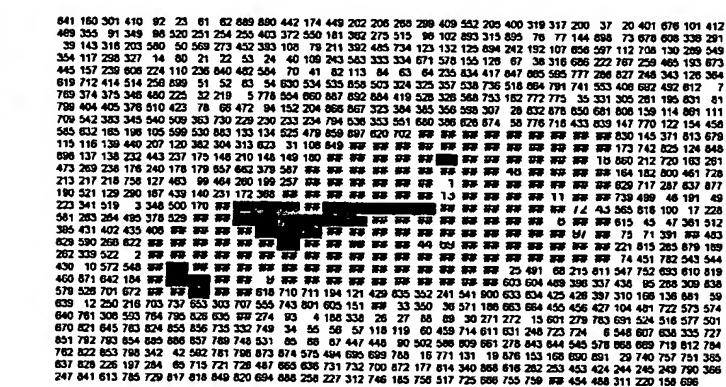
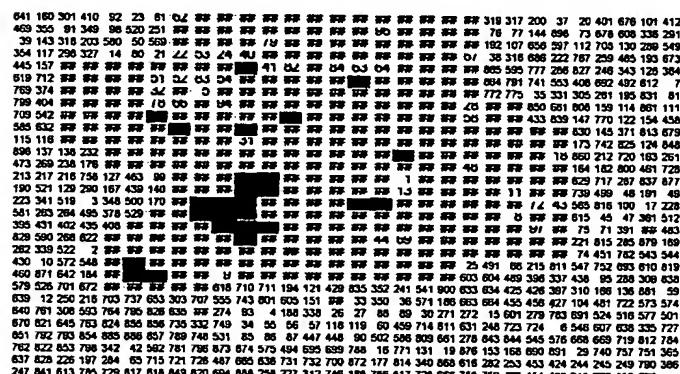
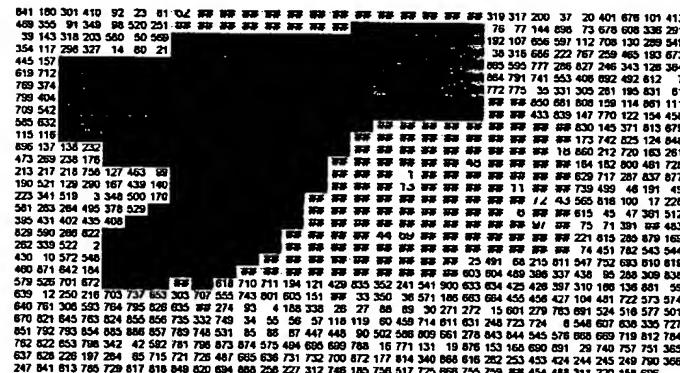
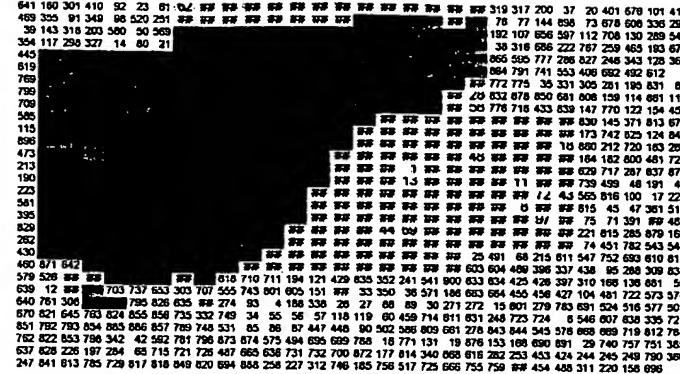
Increasing cytokinin concentration in the growth medium

Increasing auxin concentration in the growth medium

G. 3A

— 1 —

Increasing cytokinin concentration in the growth medium



Increasing auxin concentration in the growth medium

Microarray data after rearrangement

FIG. 3B

Increasing cytokinin concentration in the growth medium

Increasing auxin concentration in the growth medium

FIG. 3C

641	100	301	410	62	23	81	64	88P	319	217	200	37	30	401	675	101	4											
485	355	91	349	800	520	251	82	88P	76	77	144	226	73	678	604	288	5											
38	143	318	203	560	60	560														192	107	555	517	112	708	130	288	5
354	117	298	327	14	80	21														38	316	606	222	767	259	482	187	246
445																			885	565	777	226	827	246	343	128	34	
618																			604	704	741	563	406	692	492	193	851	
768																			772	775	35	331	305	301	122	154	54	
759																			88P	88P	650	651	808	159	111	861	1	
700																			88P	833	833	147	770	122	154	54	1	
585																			88P									
115																			830	145	371	813	67					
898																			173	742	822	124	84					
473																			212	723	123	26	154					
213																			620	717	749	178	301	361	17	22		
190																			730	449	449	191						
223																			565	816	100	17						
581																			815	45	285	47	361	51				
365																			73	71	391	89	17					
203																			221	815	285	478	170					
262																			74	451	723	543	543					
430																			25	491	68	215	811	547	542	623	610	
460	871	842	184																803	804	804	306	307	438	85	284	308	
578	592	701	672																803	804	804	306	307	438	85	284	308	
639	12	250	218	703	737	653	303	707	565	743	801	805	101	88P	33	350	36	571	802	804	455	455	455	130	181	81		
840	761	308	593	704	705	826	635	838	274	93	4	188	338	28	27	58	88	30	271	273	15	801	279	763	901	524		
670	821	645	703	824	855	858	302	749	34	55	55	57	118	119	80	459	714	811	631	248	723	724	54	546	807	833	326	
851	792	793	854	858	858	857	789	748	531	85	88	88	87	447	448	90	562	568	609	801	278	843	844	845	578	606	813	
702	622	853	798	342	42	592	781	796	673	874	575	494	495	699	788	17	811	171	131	158	163	163	169	801	284	28		
247	841	613	785	729	817	818	844	820	864	888	258	227	312	746	185	753	517	723	866	755	755	454	454	488	311	226	158	

Microarray data after rearrangement

increasing cytokinin concentration in the growth medium

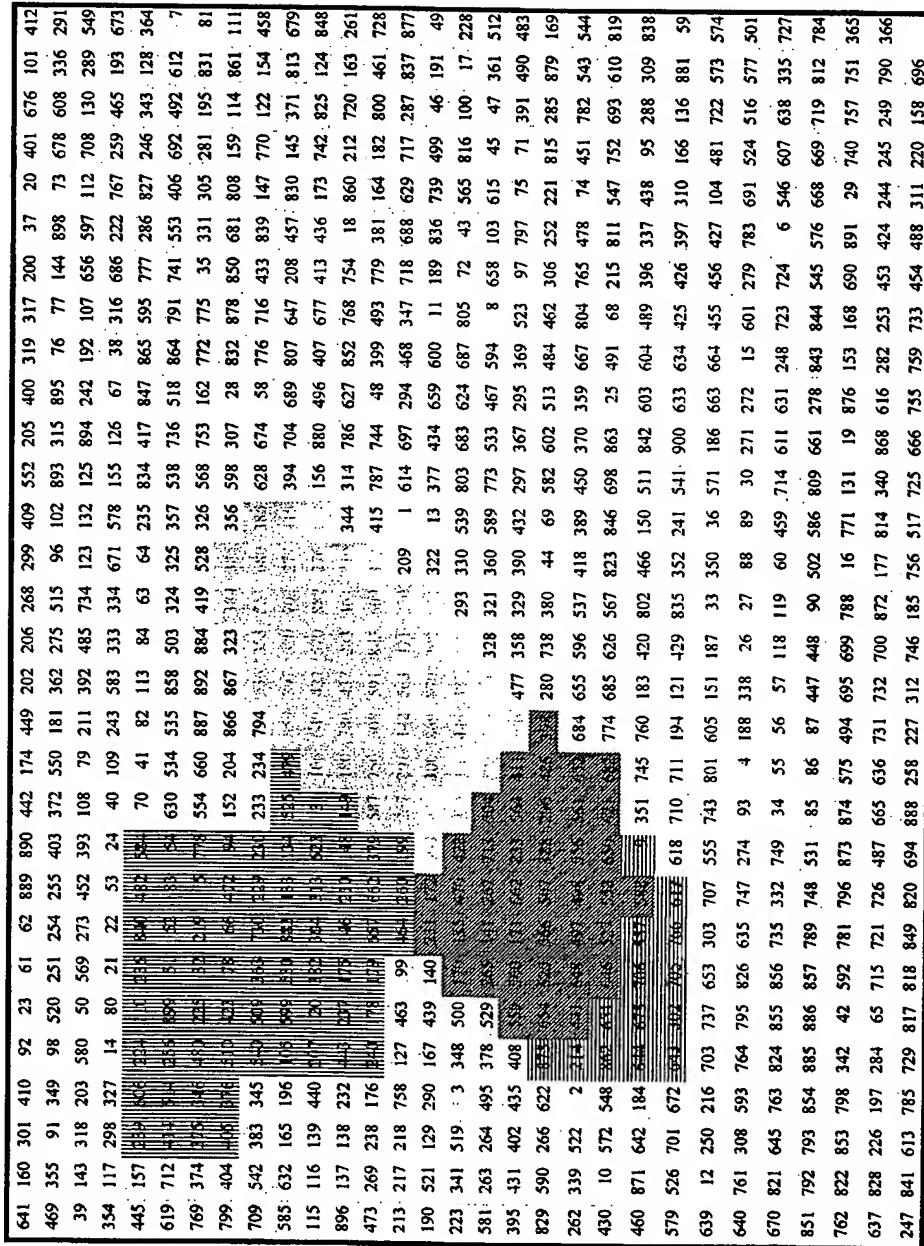
Increasing auxin concentration in the growth medium

Microarray data after rearrangement

FIG. 3D

FIG. 4

The rearrangement indicates the function of the different genes



Increasing cytokinin concentration in the growth medium ↑

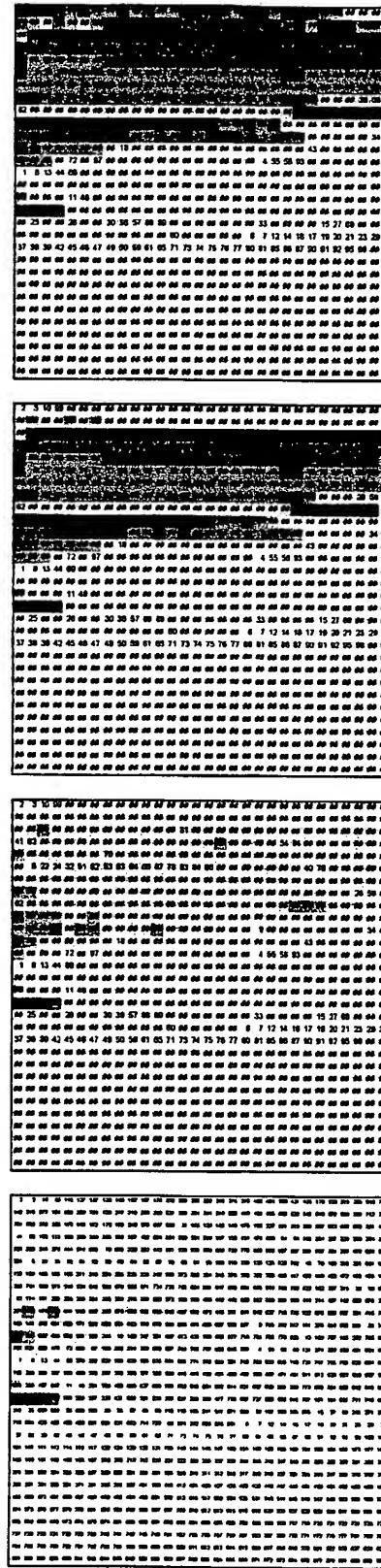


FIG. 5A

Increasing auxin concentration in the growth medium →

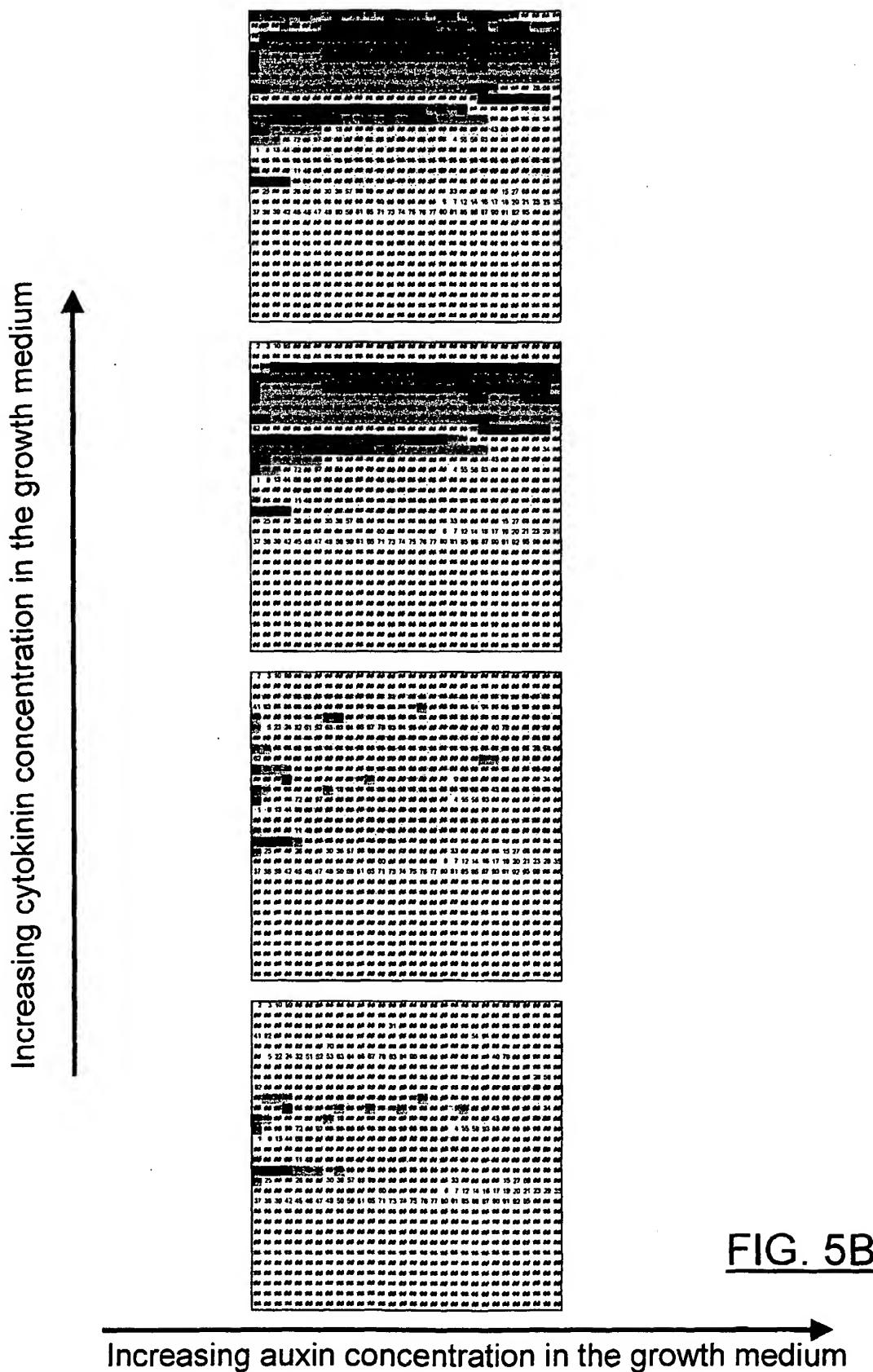
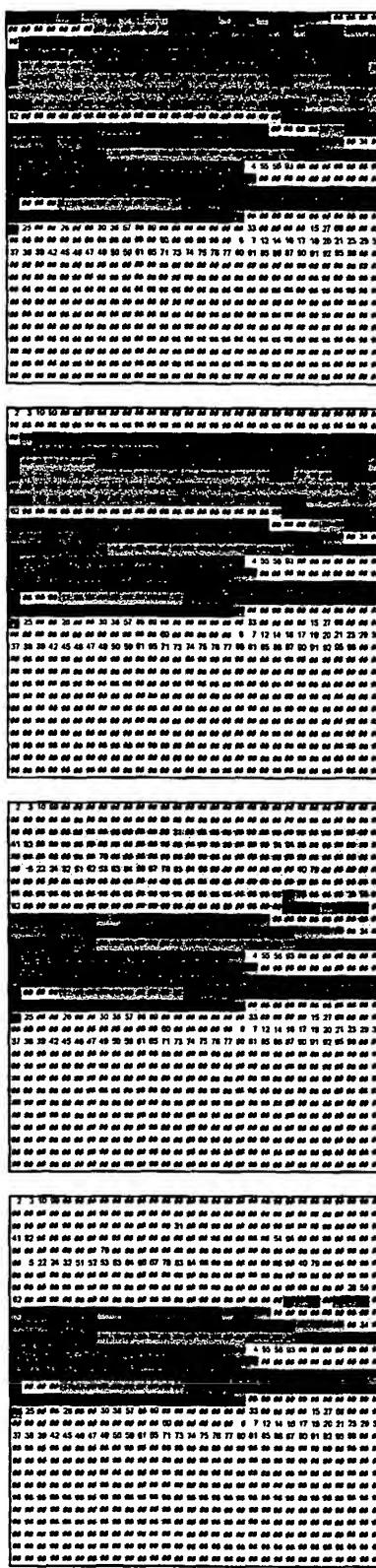


FIG. 5B

Increasing cytokinin concentration in the growth medium



Increasing auxin concentration in the growth medium

FIG. 5C

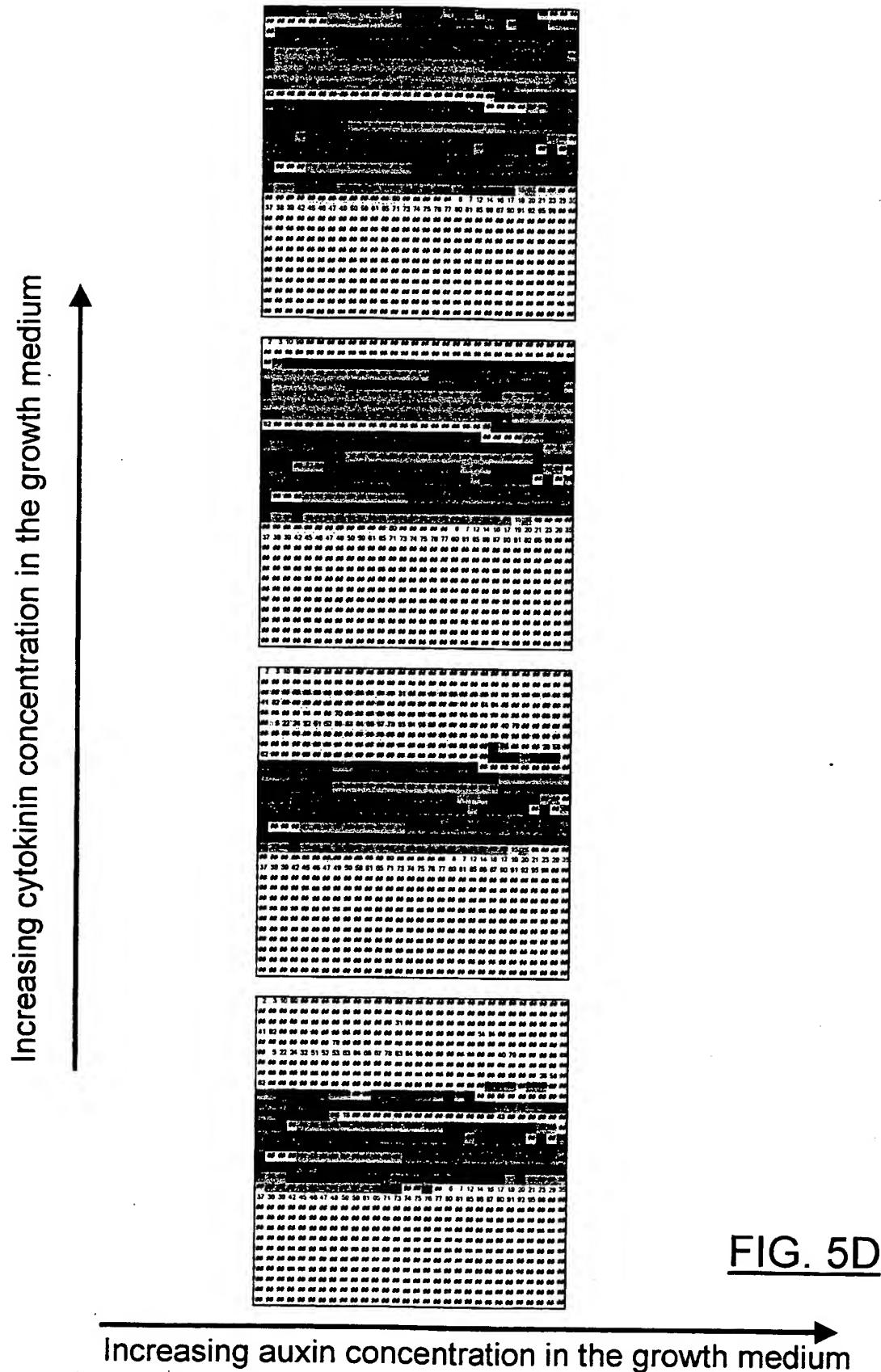


FIG. 5D

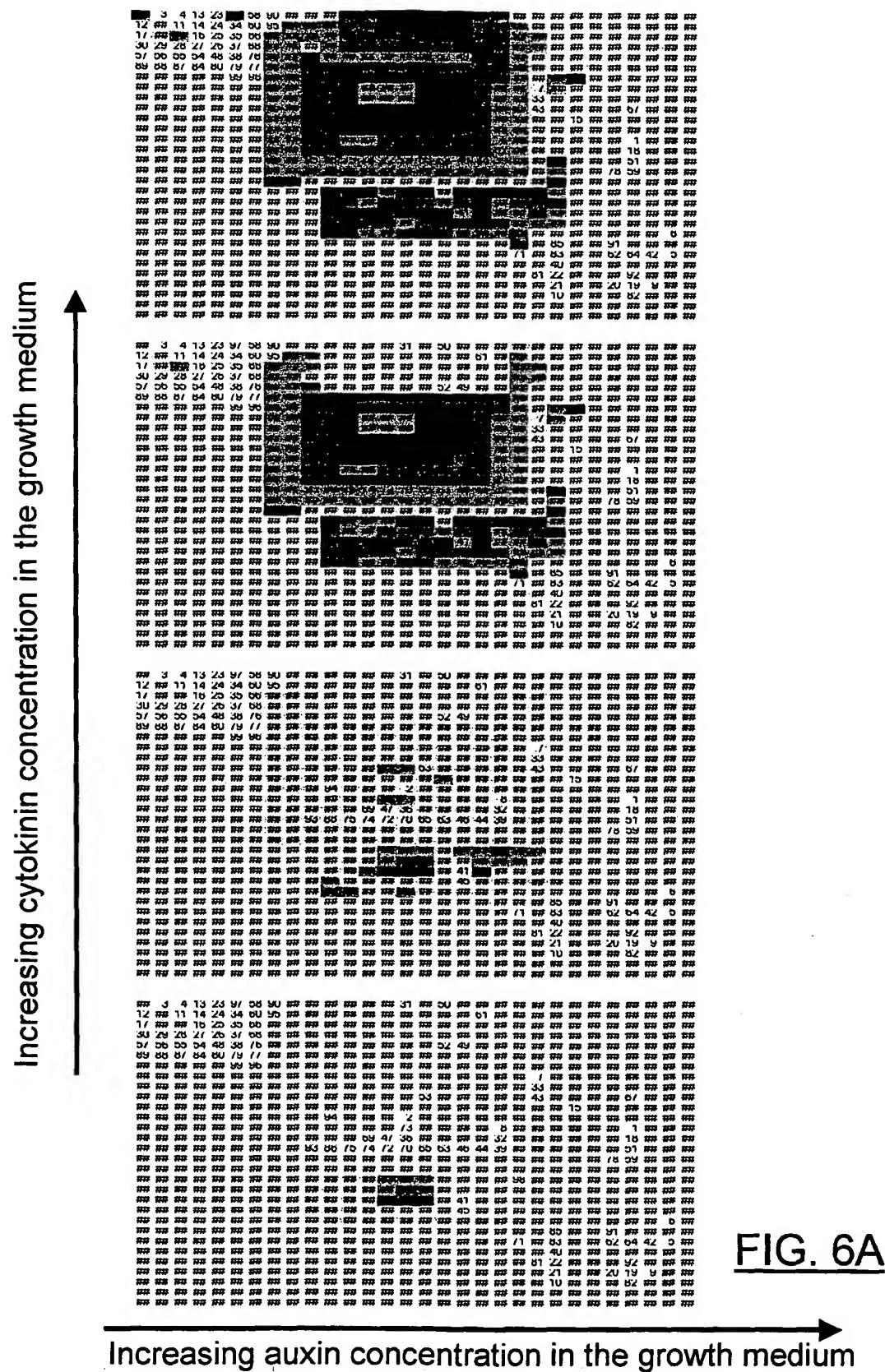


FIG. 6A

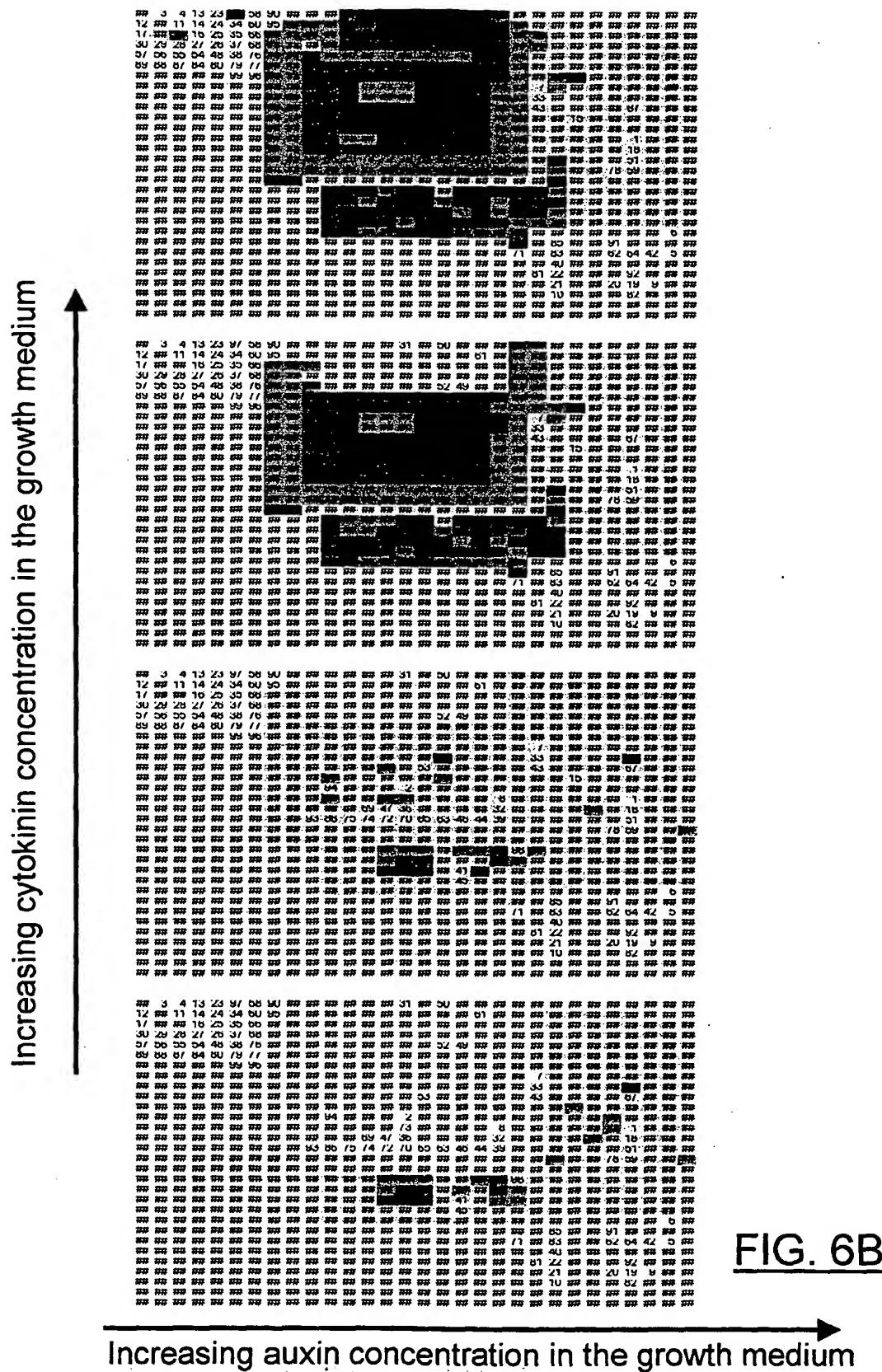


FIG. 6B

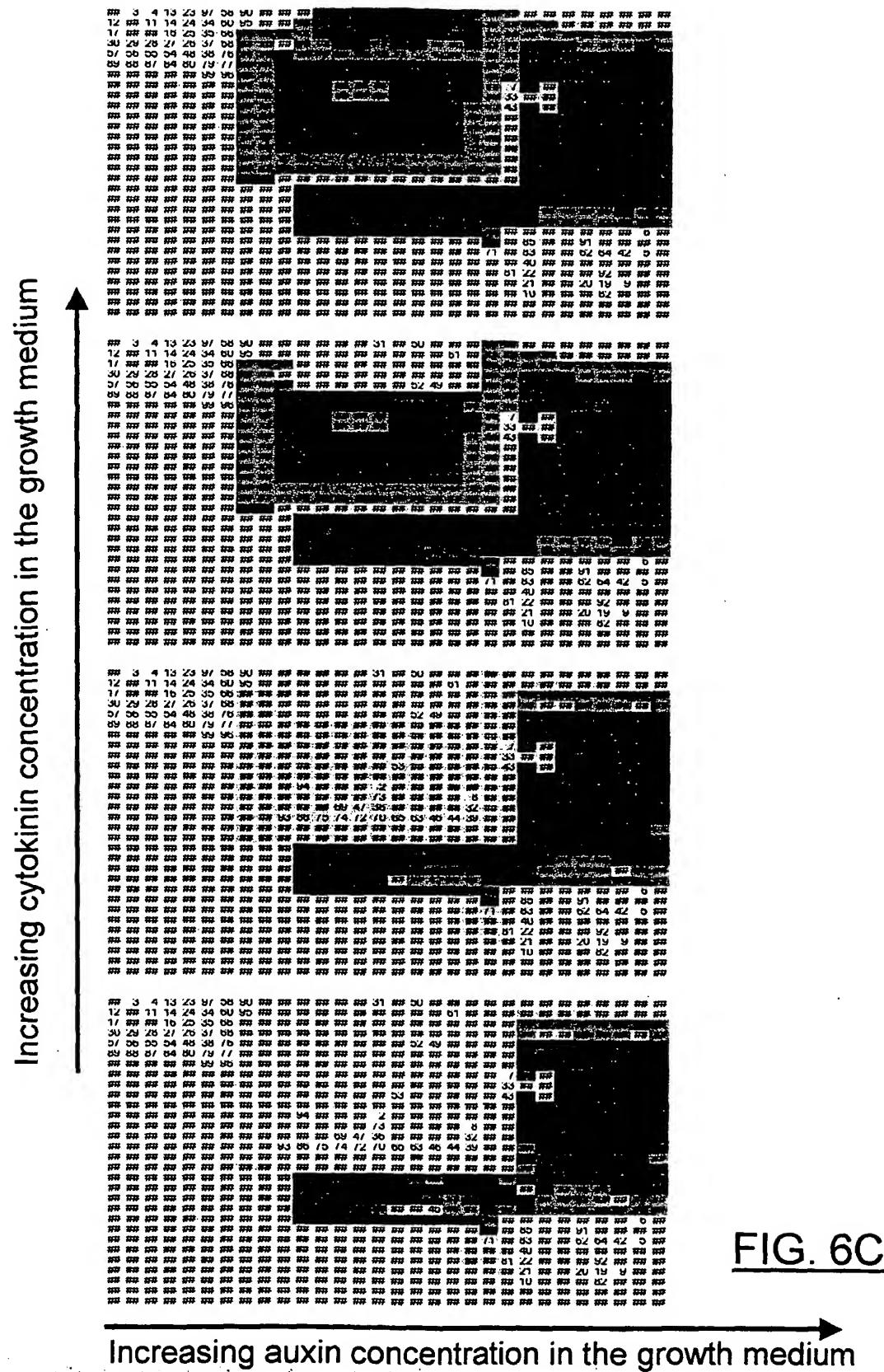


FIG. 6C

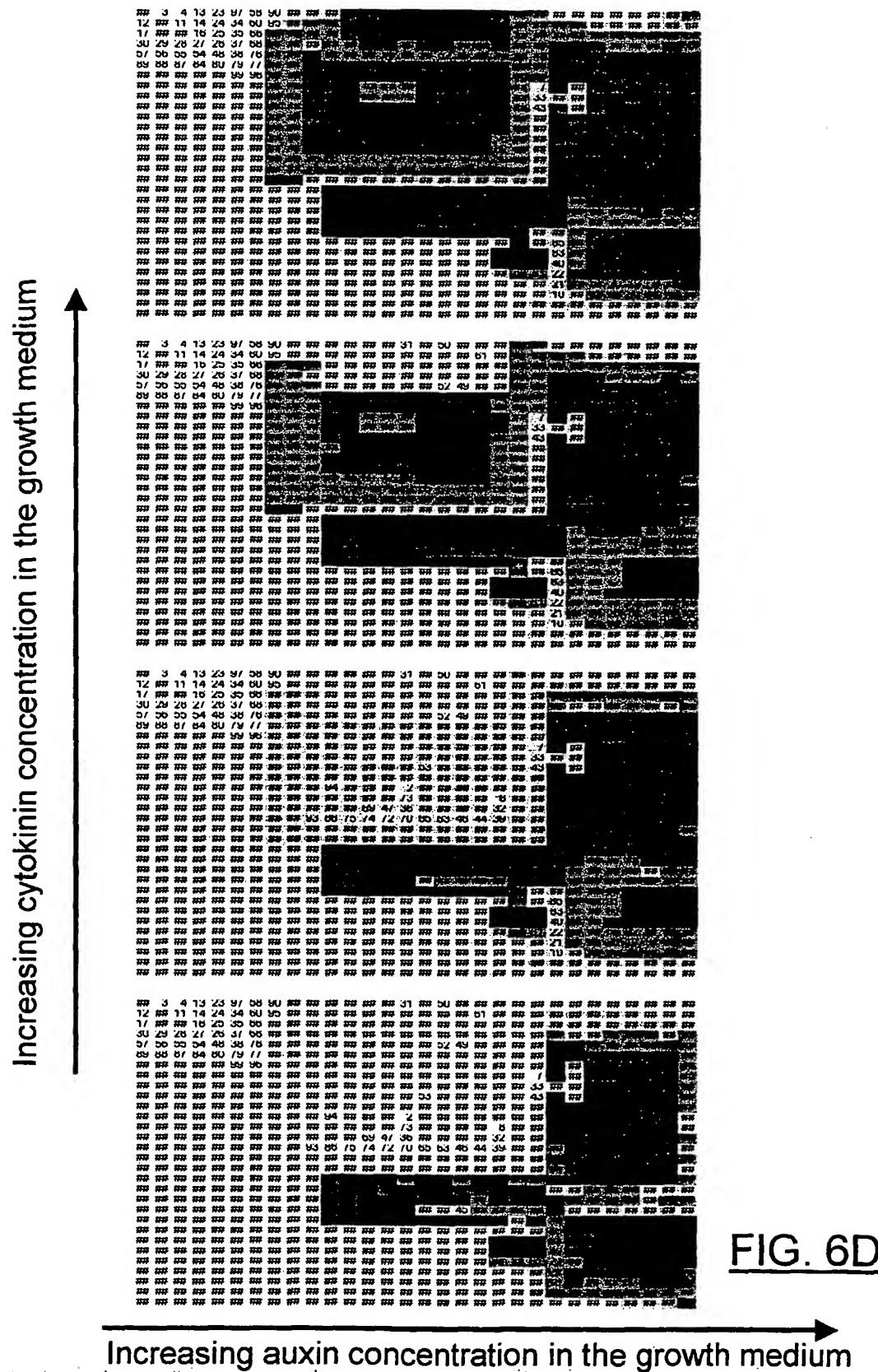


FIG. 6D

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number
WO 01/027809 A3

(51) International Patent Classification⁷: **G06F 17/30**

(21) International Application Number: **PCT/NL00/00742**

(22) International Filing Date: 16 October 2000 (16.10.2000)

(25) Filing Language: Dutch

(26) Publication Language: English

(30) Priority Data:
1013297 15 October 1999 (15.10.1999) NL

(71) Applicant (*for all designated States except US*): **PLANT RESEARCH INTERNATIONAL B.V. [NL/NL]; Droevedaalsesteeg 1, NL-6708 PB Wageningen (NL).**

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **VAN DER KRIEKEN, Wilhelmus, Maria [NL/NL]; Laantje van Anton Pieck 2, NL-6708 RE Wageningen (NL). KODDE, Jan [NL/NL]; Van Doesburglaan 28, NL-6708 MC Wageningen (NL).**

(74) Agents: **LAND, Addick, Adrianus, Gosling et al.; Arnold & Siedsma, Sweelinckplein 1, NL-2517 GK Den Haag (NL).**

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

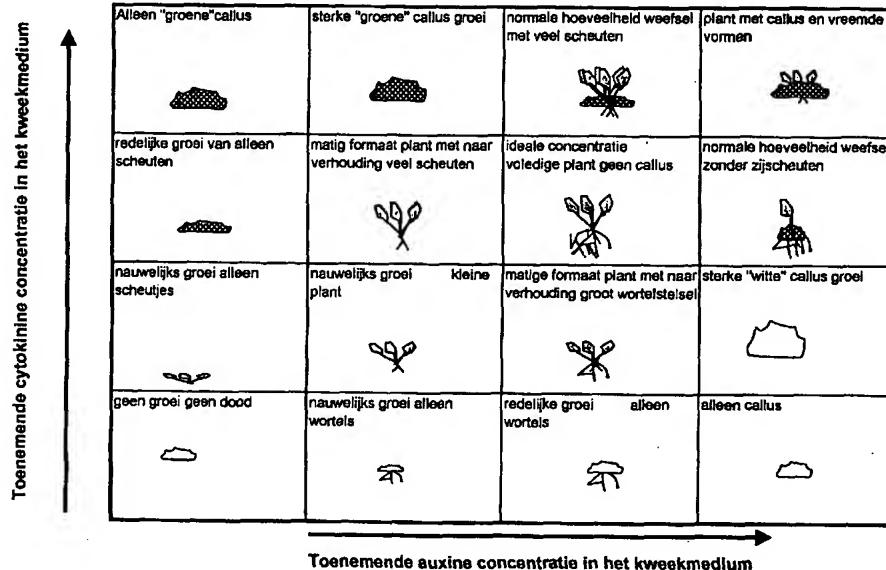
Published:

— with international search report

[Continued on next page]

(54) Title: **VISUALIZING RELATIONS IN DATA SETS**

Regeneratie experiment met plantaardig materiaal, waarbij twee parameters varieren



WO 01/027809 A3

(57) Abstract: The present invention relates to a method for comparing and/or analyzing data files, such as for instance are obtained in one or more array forms or in a matrix form, wherein: the data and relations from the arrays or matrices with mutual correlations are placed via rearrangement in a virtual matrix, whereby these mutual relations or correlations can be made easily visible for a user, for instance on a screen.



(88) Date of publication of the international search report:
12 September 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/NL 00/00742

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>EISEN M B ET AL: "Cluster analysis and display of genome-wide expression patterns" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, US, vol. 95, December 1998 (1998-12), pages 14863-14868, XP002140966 ISSN: 0027-8424 page 14863, left-hand column, line 1 -page 14864, right-hand column, line 44</p> <p style="text-align: center;">-/-</p>	1-10

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 May 2002	13/06/2002
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Authorized officer Fournier, C

INTERNATIONAL SEARCH REPORT

International Application No
PCT/NL 00/00742

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WEINSTEIN JOHN N ET AL: "An information-intensive approach to the molecular pharmacology of cancer." SCIENCE (WASHINGTON D C), vol. 275, no. 5298, 1997, pages 343-349, XP002199806 ISSN: 0036-8075 page 344, left-hand column, line 18 -page 345, right-hand column, line 49; figures 1,2 -----	1-10